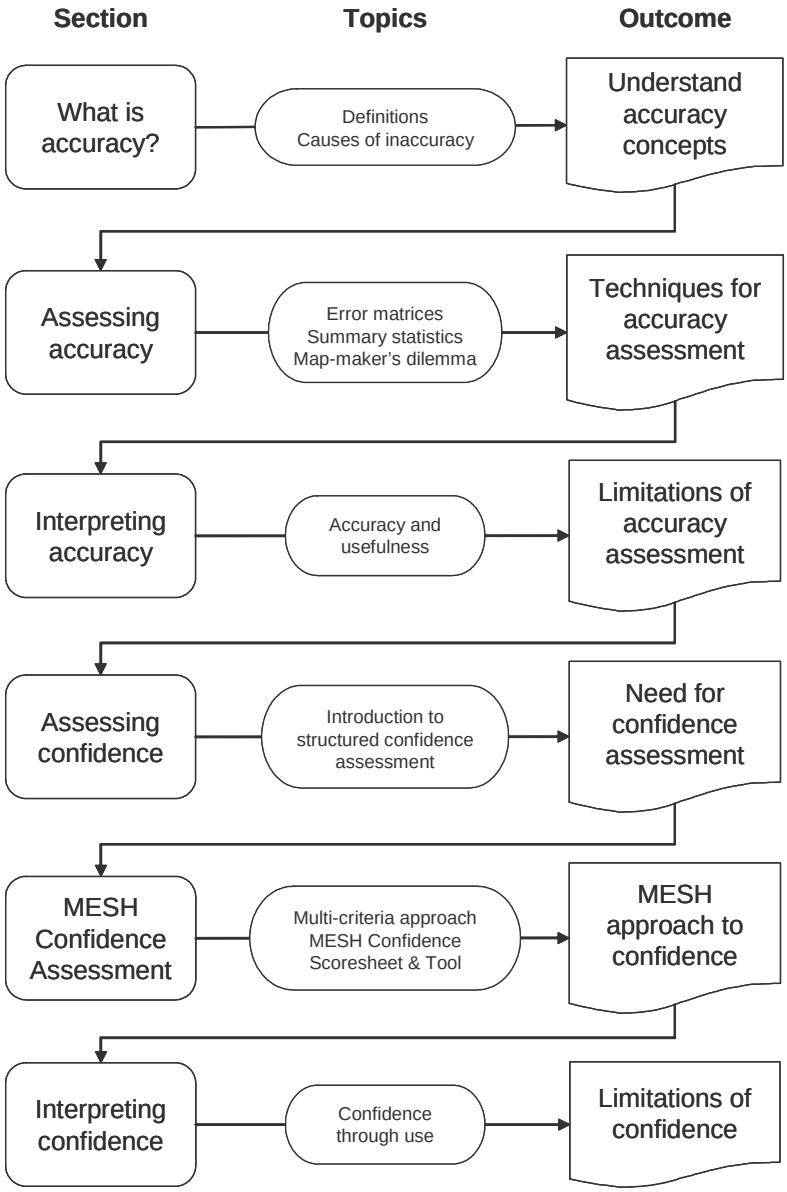
Title:	**MESH Guide: How good is my map?**
Author(s):	Bob Foster-Smith (Envision), Natalie Coltman (JNCC) and Fiona Fitzpatrick (Marine Institute)
Document owner:	Natalie Coltman (Natalie.Coltman@jncc.gov.uk)
Reviewed by:	Roger Coggan (Cefas), Jacques Populus (Ifremer), Dave Long (BGS), Jon Davies (JNCC), David Connor (JNCC)
Workgroup:	
MESH action:	Action 2
Version:	Version 1
Date published:	August 2007
File name:	GMHM5 How good is my map.pdf
Language:	English
Number of pages:	35
Summary:	The *MESH Guide to habitat mapping* aims to provide a methodological framework for marine habitat mapping so that future mapping studies will produce high quality data and maps which are inter-compatible and their outputs can be assimilated into common, harmonised maps. It will help to make habitat maps more compatible by illustrating tried and tested standards and procedures in a step-by-step manner. This document describes the issues relating to accuracy and confidence in maps. The section contrasts the mathematical approach to map accuracy measurement to the user-based assessment of confidence and introduces the MESH Confidence Tool.
Reference/citation:	Foster-Smith, R., Coltman, N. & Fitzpatrick, F. 2007. How good is my map? In: *MESH Guide to Habitat Mapping*, MESH Project, 2007, JNCC, Peterborough. Available online at: (http://www.searchmesh.net/default.aspx?page=1900)
Keywords:	
Bookmarks:	
Related information:	This document is a printable version of the MESH Guide website: http://www.searchmesh.net/default.aspx?page=1659

Close

How good is my map?

Bob Foster-Smith, Natalie Coltman & Fiona Fitzpatrick

This section of the MESH guidance aims to give an awareness of the issues relating to accuracy and confidence in maps. The section contrasts the mathematical approach to map accuracy measurement to the user-based assessment of confidence and introduces the *MESH Confidence Tool*.



A large number of terms can be used that could be applied to a map: accuracy, confidence, precision, value, usefulness, reliability and so on. Some of these terms are very subjective; others imply some independent measure we could use in any assessment. What do these terms mean and how should we use them? Perhaps a good starting point is to discuss accuracy and error, confidence and uncertainty.

Accuracy is a measure of the predictive power of a map to represent the world as measured against reality. If a map predicts a habitat at point X and it is found to be there, the map is right. If it is not, then the map is wrong.

Confidence is an assessment of the reliability of a map given its purpose. It is much more subjective and may involve a judgement of the relative importance of many contributing factors, such as level of information, how near the map is to reality ('near misses?'), how relevant to the purpose and so on.

If you are using a map, how confident can you be that the information it contains reliable? If you have commissioned a survey, how certain are you that the resulting habitat map matches your expectations? If you have produced a map, how can you best convey to others its accuracy and its limitations?

An assessment of the usefulness of a map will depend upon the intended purpose and application of the map. A map may be very useful as a broad overview, but of little use for an application where accuracy at a detailed level is required. What sort of information should accompany maps to alert people as to their legitimate use? These are very difficult questions to answer and in this section of the MESH guidance, we discuss issues that should be considered so that map users have a realistic expectation of maps without undermining the valuable contribution habitat maps undoubtedly make to marine spatial planning. The MESH program has covered an extremely wide range of mapping scales and this present section gives examples that illustrates this diversity and gives guidance on how accuracy and confidence can be assessed.

Habitat maps range in scale and detail from the small scale, broad-brush to the large scale and highly detailed. Whatever the scale, maps will vary considerably as to their reliability. *What is habitat mapping?* gives more discussion on scale and map purpose. However, the way in which we assess the value of a map may be very different depending upon scale-related issues. Broad scale maps covering very large areas are created from multiple sources and the resulting habitat map must be judged by the credibility of these sources of data together with a proper evaluation of the process that has combined these data. At the other extreme, a large scale, single survey of a small area might be judged by the precision of the survey data and the accuracy of the habitat map. Often an accuracy assessment has not been made by the original surveyors and it is left to the judgement of the user to determine the appropriate level of confidence they will place in the map. In this section we suggest ways both for map makers to measure the accuracy of their maps, and for map users to objectively judge the confidence they will place in the map so that map usage can be justified. This last point is particularly important since maps are often created for one purpose but used by others for a different application.

The section starts with a general discussion on accuracy and confidence. This is followed by a discussion of the problems of accuracy measurement and the interpretation of accuracy assessments: what does it mean if you map is inaccurate? This leads on to discuss how much confidence a user should have in a map, with a description of the MESH approach to confidence assessment as an example. The MESH partners have developed an easy to use multi-criteria system for assessing confidence of seabed habitat maps. The approach was developed to facilitate the determination of confidence in habitat maps displayed on the MESH webGIS (http://www.searchmesh.net/webGIS). Lastly, maps may best be judged against the success of their use. Were they found to be useful? Were the predictions of habitat

distribution they contained accurate enough for the purposes to which they were put? The final section of this section briefly addresses these questions.

Links to websites:

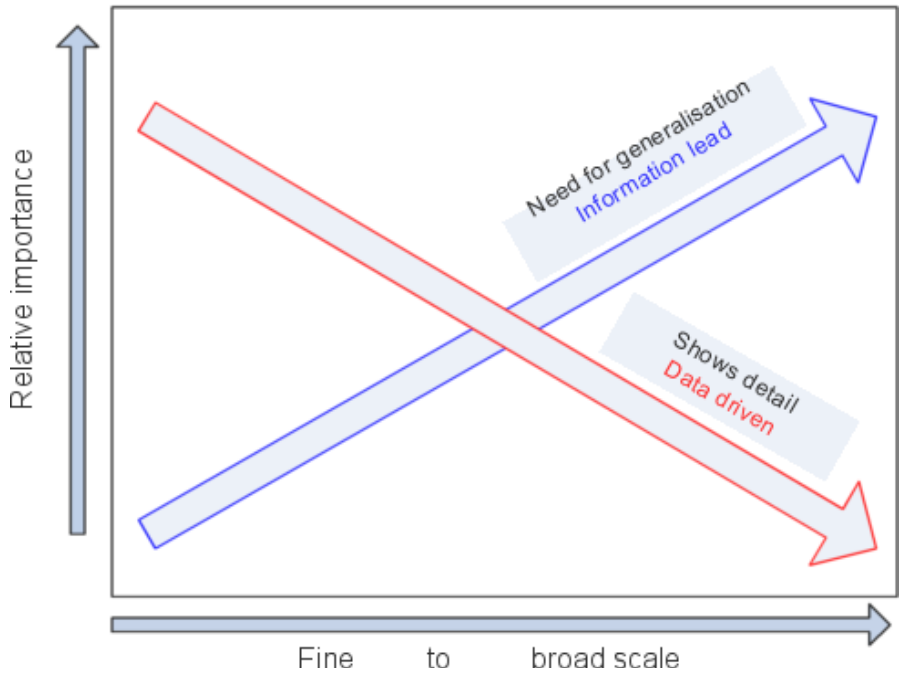
http://www.searchmesh.net/webGIS

What is accuracy?

Accuracy as applied to habitat mapping is a measure of the predictive power of a map to represent the world as measured against reality and error is a measure of the departure of a map from reality. It is a mathematical measure based on 'hits and misses' (successful predictions and erroneous predictions). Error is a measure of inaccuracy. Note that this definition of accuracy is focused on the correct prediction of a habitat class at a particular point (in a vector map) or pixel (of a raster map). In other words, there are two elements to accuracy - the right CLASS at the right PLACE. This definition is often termed classification accuracy ('have the data at point X been correctly classified?'). Clearly, there is a positional element to this accuracy. For example, are boundaries between adjacent habitats accurately located? This could be restated as 'does the change in predicted habitats accurately mark the boundary between them in reality?'

Accuracy could be used as one of the criteria for assessing confidence. However, a strict mathematical measure of accuracy could be misleading, especially if two or more maps are being compared. For example, one map might class habitats in an area as either rocky or sandy and map these two classes with a high level of accuracy. Another might show each of these habitats as a patchwork of different types of rocky or sandy habitats. The second is likely to be far less accurate, but contain more useful information *allowing for a certain level of error*. The italicised phrase stresses the important point that some user-judgement has entered the assessment to make allowances for the lower accuracy. Thus, a user may have more confidence in the information contained in the second map despite its lower accuracy. The problem is that although many of the accuracy measures are mathematically sound, they still do not address the main issue of the overall confidence with which maps should be regarded. The same measure applied to different maps may give an erroneous impression of their relative 'success'.

Indeed, there is often a trade off between information content and accuracy of a map: A map showing a large number of classes on a particular theme contains more information than one with a small number of classes. However, the error associated with the predicted distribution of the former might be quite high.



A schematic showing the changing relative importance of generalisation versus detail as the scale of a map changes

What do we mean by accuracy and inaccuracy?

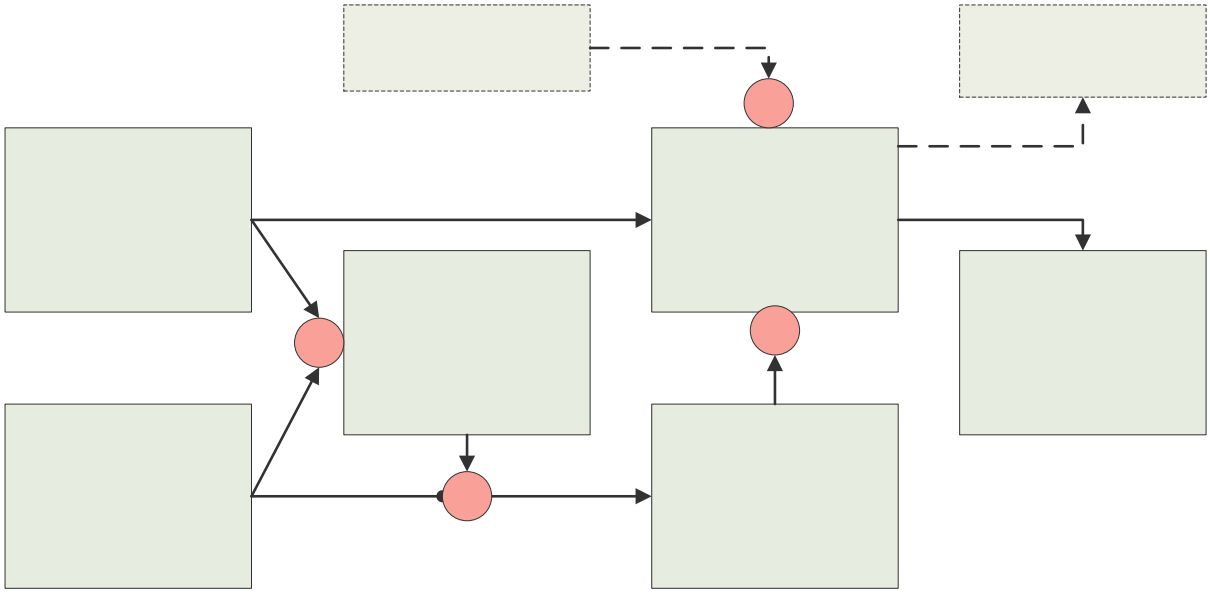
It is worth following the stages in map production to trace the derivation of accuracy since this underlines some important principles about mapping and accuracy that will help in all the discussions that follow.

Step 1: the derivation of the relationships between the ground truth data and the remotely sensed data. Since these relationships will not be perfect, there will be a margin of variability around the general trend in the predicted relationships. This often termed a margin of 'error', but is better thought of as variability.

Step 2: these relationships are applied to the whole remotely sensed dataset. Steps 1 & 2 need to be done since only a very small proportion of the area is sampled and the habitat map is, in fact, predicted on the basis of the relationships. However, since there is variability in the relationships, there will obviously be departures of observations from the predicted.

Step 3: variability is only to be expected, but manifests itself as an 'error' that measures the magnitude of the variation from the trend. If the ground truth data are used for this measurement, then the accuracy is termed internal.

Step 4: If an external ground validation dataset is used that is independent of the modelling process, the accuracy is termed external.



Variability and error

If the predicted values were continuous measurements (such as silt content in relation to depth) then the departures of observations from the predicted would be seen as variability. However, habitat maps predict habitat classes which are categorical data. Variability is, unfortunately, expressed as different classes from the one predicted and this is easily seen as an 'error'. What do we mean by a successful prediction (or conversely, a prediction error) for categorical maps? Look at some of the consequences of variability; faithfulness and exclusivity:

Faithfulness

Would we expect a habitat to occur only within limited predicted environmental conditions (i.e., it is faithful to those conditions) or also have a possibility of occurring elsewhere? Clearly, the greater the variability in the habitat/remotely sensed parameter, the less faithful is the relationship.

Exclusivity

Would we expect only the predicted habitat (i.e., we expect the habitat/environmental relationship to be exclusive) or would we expect there to be a chance of other habitats as well? Exclusivity is often forgotten when deriving relationships between a dependent habitat and an environmental parameter. The relationship might be very strong but that same relationship might hold equally well for another dependent factor (habitat). Discrimination between the two might be difficult.

Predictive power

What is meant by predictive power? Strictly, this treats maps as a series of hypotheses (predictions of 'what' will be 'where') and the more successful the map is in its predictions, the more powerful the map is. Clearly, if the correlation between a habitat and its environment is weak, the explanatory quality of statistical test (its success in explaining trends *within* the data) will be poor. However, just because the quality of the test may be high (e.g. there is a strong correlation between an explanatory variable, such as silt, and a particular habitat) this does not necessarily mean that the statistical model will be adequate to predict a wide range of habitats

Ground truth records (Sparse point data)

Remotely sensed (Full coverage & detailed data)

because of the low exclusivity of these habitats with an explanatory variable (silt may correlate strongly with many different habitats – but which one will we find in silt?).

In addition, the relationship may hold up well in some parts of the map but not in others. Perhaps there has been a spatial sampling bias that coincides with spatial trends in explanatory factors that have not been modelled so that the map may be found to have a low predictive power overall.

Spatial error

Would we consider a prediction to be false if the predicted habitat does not occur at the precise location? Or would we be satisfied if it is found within a reasonable distance? Border areas around habitats can occupy a considerable proportion of the total survey area (think of paths around a garden). This proportion grows as the habitats decrease in size (heterogeneity) and the width of the border zone (the 'path') is increased. It is also inevitable that the border zones are likely to be the places where uncertainty is highest. It follows that this is going to affect our impression of error. Heterogeneous areas are much more likely to have high errors than homogeneous areas. We must consider heterogeneity when interpreting the error measurement.

Probability

How can we best express predictions? It is common to express predictions as probabilities and likelihoods. In its weakest sense, if we say that a particular habitat is most probable at a particular location we might simply be expressing the strength of our belief based on experience and the evidence (e.g. from a visual inspection of a side scan sonar image by an expert), A statistical expression of probability is based on an analysis of the available data and ranges from close to 1 (a prediction with a very high probability) to close to zero (an extremely low probability). Maps can show the distribution of a habitat in terms of its probability of occurrence. However, these probabilities cannot be shown for all habitats on the same map because more than one habitat is likely to be predicted for every location. Most habitat maps, therefore, show only the habitats with the highest probabilities (see section Can I map uncertainty?). It is very important that we remember that for most habitat maps there are underlying, competing probabilities that are hidden from the map users. They only see the winners!

Why are maps inaccurate?

There are two main reasons why maps may not match reality very well. Firstly, we are limited in the way we represent the real world: benthic habitats are very complex and multifaceted and yet we need to reduce this complexity to a small number of habitat classes (categorical data) for mapping. Often matching observations to classes is not clear and this gives rise to ambiguity and thence to an apparent mismatch when a map is compared to observations. Secondly, the process of measurement, analysis and cartography can introduce error. A map will combine both ambiguity and error. Accuracy measurement is part of the process of determining how close to reality the mapped distribution patterns are, set against this background of ambiguity and error. Ambiguity and error, therefore, combine to create uncertainty – an assessment of the lack of confidence in a map. There is no easy solution to confidence assessment and, ultimately, the goal must be to produce maps with levels of confidence commensurate with the information required for the map's intended purpose.

Sources of uncertainty

The ideal map would be accurate with a high level of precision and contain all the information that might be required by users. For example, fine-scale ordnance survey maps might be expected to show man-made objects in their correct positions with a very small margin of error. This is not the case for benthic habitat maps! Maps show the way the map-makers see the sea floor making best use of the data available to support their viewpoint. The following provides a description of some of the main causes of uncertainty:

Measurement error of the ground truth data

The natural environment is extremely complex and we need to simplify the real world considerably for mapping. The objects we map are usually our attempt to force the highly variable nature of the world into a manageable number of categories. It is inevitable that there will be ambiguity in this process which can originate from various sources:

- Variability in the way surveyors apply a classification process to record data. The definitions of classes will be vague and many of the criteria will overlap from class to class. Error can be minimised by better definition of class attributes and standardised protocols for assigning samples to classes. Absence of clear guidelines makes it difficult and interpretation subjective.
- Real heterogeneity on the ground. Variability is complex firstly because habitat features very often are on a continuum and lie between two or more habitat types in the classification, and secondly because fine-scale heterogeneity may result in the minimum mapping unit (MMU) or pixel encompassing more that one class.
- Trying to fit observations limited by the technique used to a classification system where classes are based on more complete information. Video observations, for example, may not provide full information on infauna and the observation is classed on the basis of conspicuous fauna.

It cannot be stressed strongly enough that interpreting remotely sensed data using ground truth observations can be undermined by poor attribute measurement of the ground truth samples. This is particularly likely when the attributes are habitat classes and the analyst must decide how best to match the sample data to a classification system.

Subjective interpretation of boundaries

Many habitats are characterised by indiscrete or diffuse boundaries and are therefore subject to the interpretation or bias of the field mapper (for direct mapping) or visual interpretation of images (e.g. side scan images).

The inherent variability within and between the remote sensing systems

All remote sensing techniques have inherent variability that degrades their ability to discriminate features on the ground. Variability may also apply to distortion in video systems and the way different grabs of the same type ‘bite’ the sea floor, introducing observation error. Calibration of equipment is vital to the accuracy of the data and utilising poorly calibrated equipment will downgrade the accuracy of the final maps.

Positional errors of remote sensing, ground truthing and combined errors

The equipment we use will have limitations as to positional accuracy. Image processing requires the location of the ground truth samples on the image so that image characteristics can be associated with the ground truth classes. The combined positional errors will give rise to a locus (or 'error envelope'). Thus, even if we could be absolutely precise about the mapping units, we could not be precise about where boundaries should be. Nor, because of discrimination, could we be absolutely sure we have detected the class with absolute certainty.

Error from analysis

Error and uncertainty will also be introduced through analysis, especially given that very often target classes themselves cannot be directly detected by remote techniques and their presence is inferred via statistical links to infauna or other observed variables (proxy maps, surrogacy). There can be many stages involved in image processing from data editing through to statistical analysis and modelling. However, the route followed by an analyst may be hard to replicate by another person since there are many possible pathways, each with different parameters that must be set. It is hoped that analysis is robust, but there is always the possibility that the interpretation is sensitive to apparently trivial parameter settings.

Error from sampling bias

Not every point in a map is validated. Maps are based on some form of sampling strategy and these data are extrapolated to the whole area using assumptions about the statistical relationship between the samples and the 'population' from which they are drawn. Wherever there is sampling there will be bias and problems of under-sampling. This is especially true for geographic systems where the uniqueness of location makes sampling strategy difficult.

Cartographic error

There is a limit to what a map can show (detail, number of classes and resolution) and maps generalise information to a greater or lesser extent. The ability to show detail in a map is determined by its scale. A scale of 1:2,000 will illustrate much finer points of data than a smaller scale map of 1:200,000. Scale restricts type, quantity, and quality of data. Enlarging a small scale map does not increase its level of accuracy or detail

Errors can multiply!

Map-makers should provide information that allows others to assess likely margins of error. This is straightforward with some measurements, for example positions can be given along with their margins of error. It is not so easy to measure the variability in the way workers have assigned habitat classes to ground truth samples or the subjective nature of visually drawing boundaries around features seen on a remotely sensed image. Even automatic classifiers (e.g., texture analysis, supervised classification) assume that the ground truth data have been accurately categorised, or accurately partitioned between classes in the case of fuzzy classification. Despite these important reservations, can accuracy measurement be used to compare one map with another or provide a universal yardstick for measuring performance?

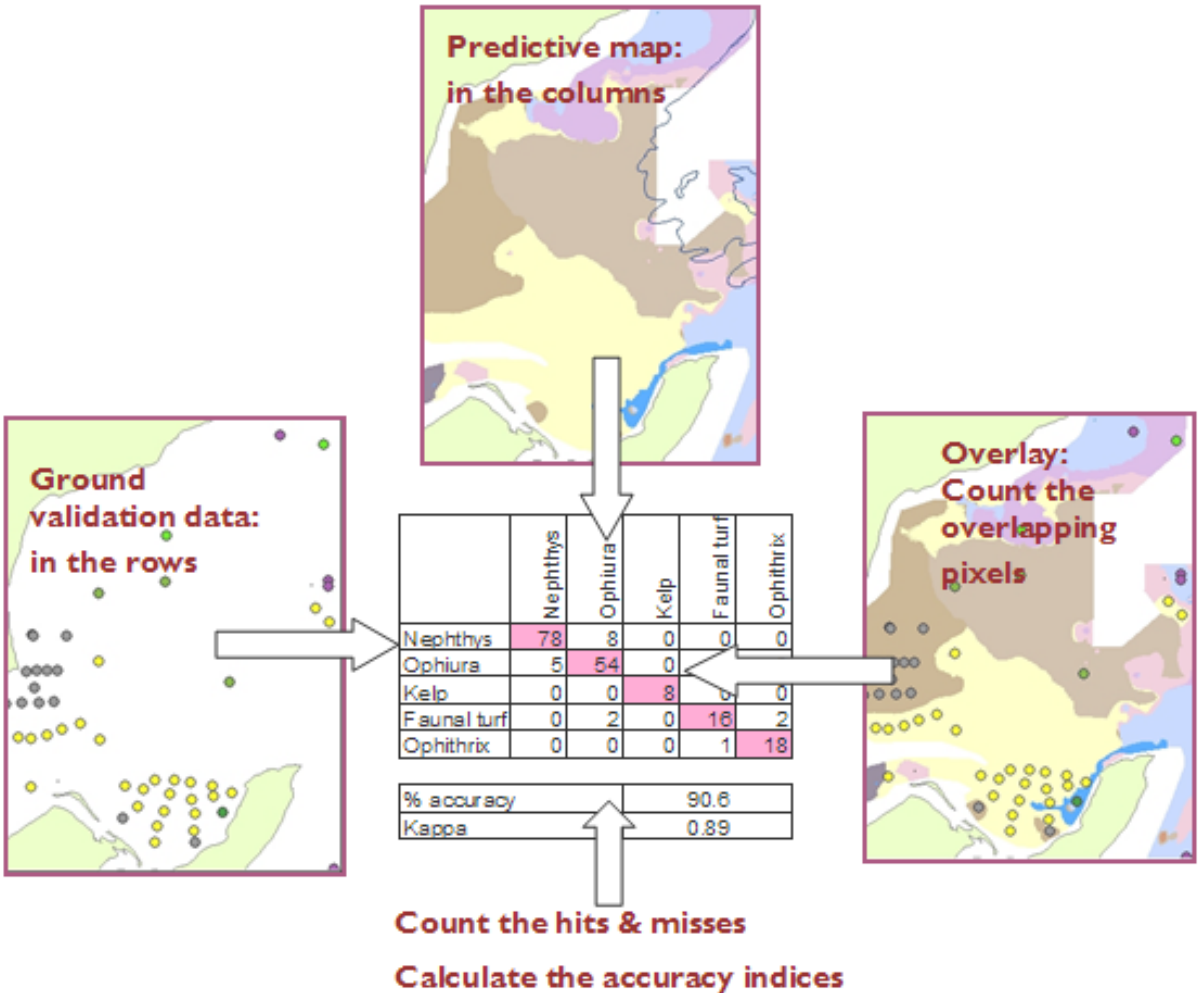
Some mapping involves only a single step combining remotely sensed and ground truth data. However, other mapping involves a complex series of steps: The remote data may be first interpreted as a sediment parameter map (e.g., silt fraction), slope, topographic feature and so on. These are then used as proxies for the habitat,

Authors: Bob Foster-Smith, Natalie Coltman, Fiona Fitzpatrick

requiring statistical relationships to be established between habitat and these derived proxy maps. Each stage can introduce its own errors that combine and grow at each successive stage. These errors can be modelled. However, a simple hit-rate comparison of the final predictive map with a ground truth habitat point dataset (or better still, an external ground validation dataset) can circumvent this chain of error estimation if an empirical measure of accuracy is sufficient and no analysis of the relative sources of error is required. Methods to assess map accuracy are discussed in the next section.

How can I assess the accuracy of my map?

Accuracy has been introduced as a mathematical measure based on the predictive power of a map to correctly predict the habitat for a particular point (or pixel). If class 'A' is predicted to be present at location 'X' and this is found to be the case by observation, then the map is right at that point; if it is not, then the map is wrong. This is the basis of all accuracy measures. If there are a larger proportion of wrong predictions, then the map is inaccurate and might not be regarded with much confidence. This is calculated by overlaying the ground truthing data (or better still, the ground validation sample data) over the predictive map and presenting the success of the match as and error matrix.



	Nephthys	Ophiura	Kelp	Faunal turf	Ophithrix
Nephthys	78	8	0	0	0
Ophiura	5	54	0		
Kelp	0	0	8		0
Faunal turf	0	2	0	16	2
Ophithrix	0	0	0	1	18

% accuracy	90.6
Kappa	0.89

Overlaying ground validation data with the predictive map to generate and error matrix.

The diagonal cells in the error matrix contain the percentage of each class that were correctly predicted. The cells which do not fall on the diagonal show incorrect predictions. The basic accuracy measure is the overall percentage correct. More sophisticated measures take into account the proportion that might be expected to be 'correct' purely by chance. Further methods of measuring accuracy are discussed below.

Authors: Bob Foster-Smith, Natalie Coltman, Fiona Fitzpatrick

Summary statistics for accuracy and error

In theory, it should be possible to measure the absolute accuracy of a map using statistics based on hit-rates: a pixel or polygon either matches a ground validation sample or it does not. The analyst then has to explain inaccuracies and, if possible, correct the procedure to improve accuracy (without fudging!). The basic tool for this is the analysis of an error matrix.

Error matrices are easy to construct for raster datasets because the image for the ground validation samples can be overlain onto the predictive habitat image, assuming the two images to have the same pixel size and format so that there is a pixel-on-pixel comparison. Most image processing software or GIS will calculate the matrix and standard accuracy measures.

The error matrix will be an N x N matrix where N = number of classes. The rows headings will be the ground validation classes and the columns the predictive map classes where there is overlap between the two images. The data in the matrix are the number of pixels of each ground validation class which fall in each predictive map class. The diagonal shows correspondence (correctly classified), the off-diagonal values indicates error. Errors of omission, where a habitat class was present at the location of a particular pixel but not predicted, can be read along the rows (less the diagonal cell). Errors of commission, where a habitat class was predicted to be present when, in actuality, it was not, can be read down the columns (again, less the diagonal cell).

From this basic matrix a number of summary statistics for accuracy and error can be derived:

- **Overall percentage correct:** [(sum of diagonal cells)/total cells in overlap] x 100.
- **Omission error** (for any class or group of classes): Pixels in rows minus the appropriate diagonal cell for the class or group of classes.
- **Producer's accuracy** (for any habitat class): Classes correctly predicted: Number of pixels of a class correctly predicted/total number of that class known to exist in the ground truth image.
- **Commission error**: Pixels in columns minus the appropriate diagonal cell for the class or group of classes.
- **Consumer's accuracy**: Pixels correctly classified: Number of pixels correctly predicting a habitat/total number of pixels of that class predicted in the classified image
- **Average accuracy**: Sum of producer accuracies for each class/number of classes.
- **Kappa** (and other similar statistics): A statistic that adjusts overall accuracy to account for chance agreement (used in preference to percentage correct).

	Lanice	Nephtys	Abra	Sabella discifera	Sabellaria	Reef	Sabella pavonina	Ensis	Ophiura	Modiolus	Error of omission for Sabellaria
Lanice	20	11	0	1	0	0	0	0	0	0	
Nephtys	7	232	16	4	4	0	7	0	0	0	
Abra	0	7	25	0	7	0	0	0	5	0	
Sabella discifera	0	12	0	17	7	0	0	0	0	0	
Sabellaria	0	11	0	0	125	16	0	8	7	0	0.25
Reef	7	11	0	0	38	58	0	0	0	0	
Sabella pavonina	0	1	0	0	0	0	8	0	0	0	
Ensis	0	0	0	0	0	0	0	8	0	0	
Ophiura	0	12	11	0	8	0	0	0	22	0	
Modiolus	0	0	0	0	0	0	0	0	0	21	
Error of commission for Sabellaria					0.34						

An example of an error matrix produced when ground validation samples are compared to map predictions

In the above example, the predictions that are verified by the ground validation data are in the diagonal, pink cells. To find errors of omission, read along the rows for cells other than the diagonal cell, for example, the yellow row highlights pixels that should have been classed as *Sabellaria* but were predicted to be one of the other habitats. The error is given as a proportion. For errors of commission, read down the columns for cells other than the diagonal cell, for example, the blue column highlights pixels that were classed as *Sabellaria* but were found to be one of the other habitats. The error is given as a proportion). In the above example, the percentage correct is 71% and the Kappa index is 0.68 (where 1 is a perfect match, 0 is entirely random). Note that in this case the error matrix also indicates that *Sabellaria* reefs and non-reef habitats are most likely to be confused (read along the yellow row). This might be expected because of the lack of a distinctive difference between these two habitats.

Ground truth, ground validation and the map-makers' dilemma

Ground validation and test accuracy (or external accuracy): The standard error matrix as outlined above is also termed the test accuracy in which external set of ground validation data was used to assess accuracy. It is important that this sample dataset was **not** also used for deriving the map and the dataset does comply with the definition of ground validation data. This is a test for the predictive power of a map.

Ground truthing and training accuracy (or internal accuracy): Samples also used for interpretation of the map (ground truth samples) are overlain on the derived habitat map. This is the most usual way accuracy is tested in maps because of the map maker's dilemma (see below).The map and the ground truth samples are clearly not independent and any measure usually exaggerates accuracy. Strictly, it is a measure the strength of the correlation between the ground truth data and the remotely sensed data and is a measure of the explanatory quality of the map. It is not a measure of a map's predictive power. However, when there are large numbers of ground truth data, training accuracy and test accuracy converge because there is

less likelihood of encountering conditions where the correlations between habitat and environment have not already been encountered. (However, this might not always be the case: larger numbers of ground truth samples may also result in weaker correlations if they are spread over broad scale environmental trends that are not taken into account).

Test accuracy is the best method for assessing accuracy but often difficult to perform in practice because of the map-maker's dilemma: ground truth data are hard won in most marine surveys and setting aside sufficient samples to act as a validation dataset would not leave sufficient data for ground truthing. The interpretation of the map (e.g. through supervised classification) would be seriously affected by the exclusion of these data. In other words, the benefits of test accuracy are likely to be outweighed by the decrease in classification performance. This is a serious dilemma for map makers in the marine environment. Test accuracy can be achieved by setting samples aside for validation accepting that this might have a slight detrimental effect on the maps. It has been suggested that about 20% of samples could be retained for this purpose. The selection of the samples to be retained could be done on a random basis or, probably more successfully, on a stratified random basis so that sufficient samples of each class were retained for classification.

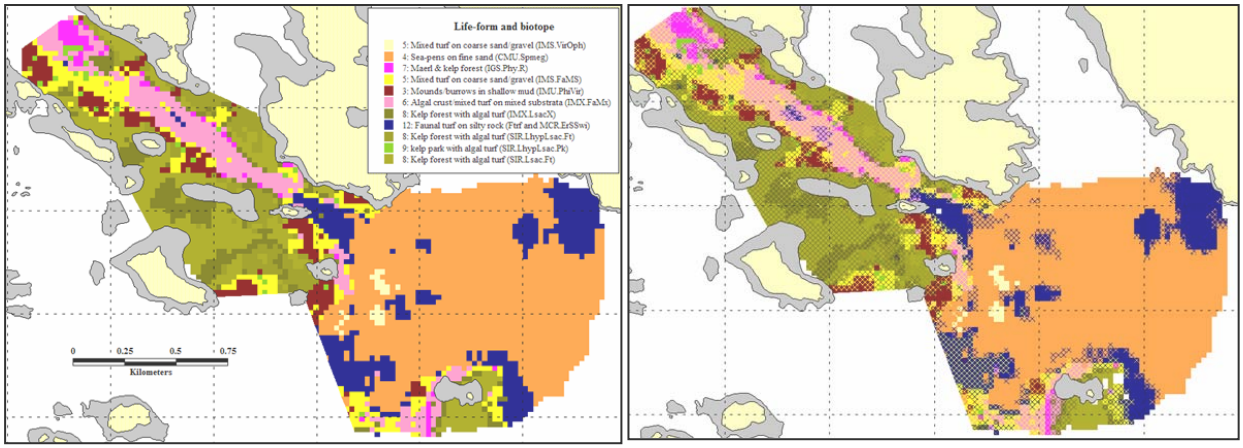
A development of this is to retain a smaller proportion for validation, but then return these samples after classification, selecting another set and repeating the classification, and so on until sufficient runs have been performed to calculate the accuracy and variability for each habitat class. However, this jackknifing technique is computationally demanding and must be regarded as a research tool rather than a standard mapping procedure.

Mapping the confusion between classes – 'fuzzy' maps

Habitat classes often show considerable overlap in the environmental conditions within which they occur, and, where mapping is based on acoustic properties, it is not always possible to distinguish habitats based on characteristics of acoustic reflectance. This is shown by the distribution of incorrect classifications in the error matrix. If the error matrix is used in this way, it is termed a 'confusion matrix'. These matrices are useful tools for measuring the overlap of classes caused through confusion between signatures.

Overlap is particularly marked, not surprisingly, between similar habitats. This situation reflects the fact that the natural environment is best represented as continua rather than discrete and separate units. Although we cannot map with such multidimensional continua, we at least have to acknowledge the 'fuzzy' boundaries between habitat classes.

This has implications for accuracy measures because instead of predictions being either right or wrong, predictions can be nearly right. Although there are ways to accommodate this fuzziness, computing this is convoluted and representing it can be confusing. This fuzziness can be demonstrated for a map by showing which habitat classes are confused and by how much through confusion matrices. Although instructive, the quantification of fuzziness is probably not easily incorporated into any assessment of accuracy or confidence.



Comparing the use of a 'hard classifer' (left) with a fuzzy classifier (right) where the fuzzy approach shows where alternative habitats may be present

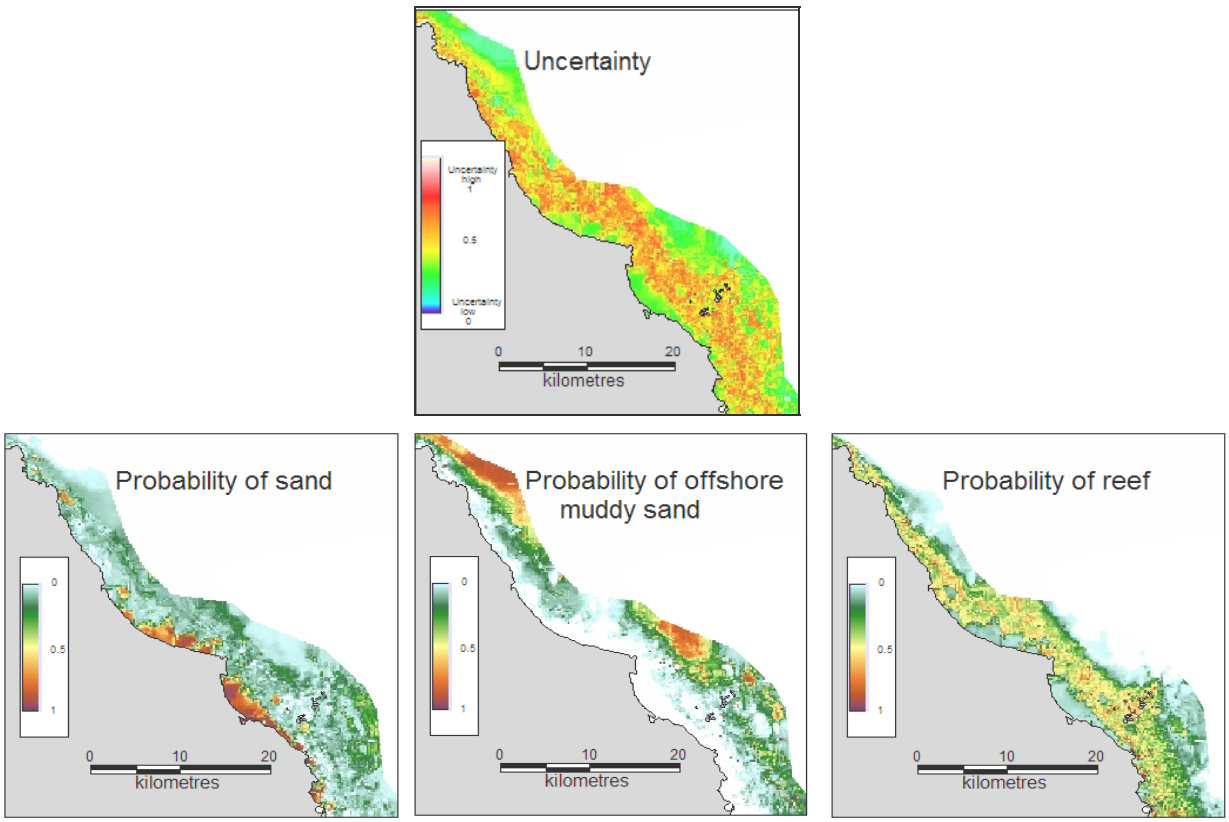
The map on the left has used a hard classifier so that only the most likely class is shown. The map on the right has second choices (where these have a high probability) as a hatched overlay. It might be a more informative map, but is it easier for a user to read? Whilst the measure of success of a map can be increased by using fuzzy procedures and also making allowances for near misses, eventually so much allowance for 'near' misses can be made that the resulting maps become unreliable.

Can I map uncertainty?

The error matrices produce an overall statistic for error for the whole map and also for each class. However, these summary statistics apply to the whole map and show no geographic variation over the map. The fuzzy maps do indicate geographic trends in uncertainty (in the above example, there is uncertainty where there is a hatch over the most probable class). Are there other ways of showing the varying degree of uncertainty over a map?

Image processing techniques provide one way of doing this. The section on supervised classification (in *How do I make a map?*) explains that when habitat class signatures are applied to the raster layers the pixels are assigned on the basis of the highest probability. This is picked from the probabilities that have been calculated for all habitat classes for each pixel. With the standard classification routine these individual probabilities are not visible. However, they can be viewed as individual layers, one for each habitat class. They can also be used to determine the level of certainty with which a pixel is classified (the more evenly the probabilities are spread between the classes, the lower the certainty). Maps can then be prepared showing certainty from 1 (one class has a probability of 1, all the others have a zero probability) to 0 (all classes have an equal probability).

Authors: Bob Foster-Smith, Natalie Coltman, Fiona Fitzpatrick



Examples of maps showing the probability of occurrence of individual habitats

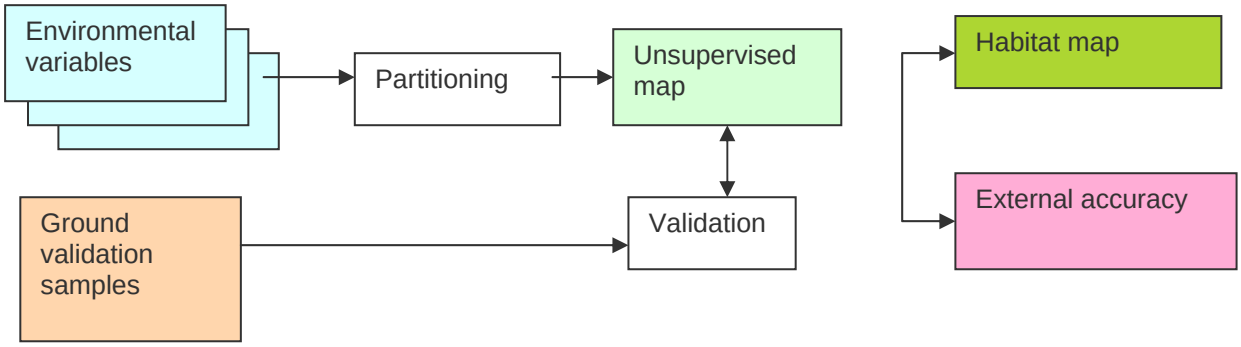
Statistical correlation techniques

There are many statistical techniques that model the dependence of variable such as percentage sand, on another variable, such a depth. These include linear regression and geostatistical techniques (Kriging-based methods). These variables can be used in their turn to model the distribution of particular species or habitats if the environmental variable is show to be a suitable surrogate for the habitat.

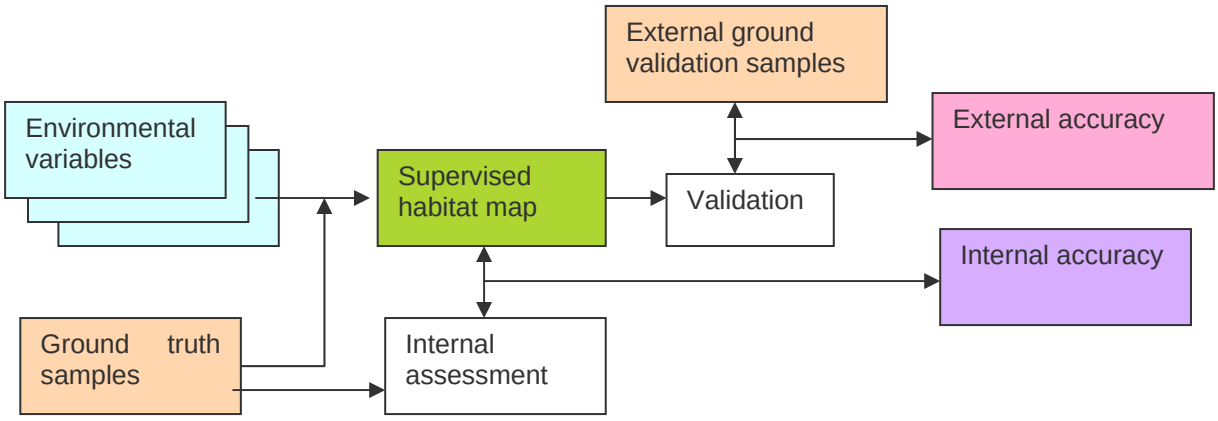
Successful models reduce the variance of the residuals of real data from the predicted values. Thus, error estimation (the converse of accuracy) is built into the modelling process. These techniques are appropriate for deriving the best distribution map of biologically important environmental variables as inputs into habitat models. Although they are important for assessing the performance of models, they are not directly applicable to categorical habitat maps. As such, these techniques are more appropriately discussed in the section on modelling (in *How do I make a map?*).

Partitioning techniques

There are many automated techniques that take one or more variables and partition an area up on the basis of distinctive combinations of characteristics (usually multivariate techniques). The next stage in the classification process is to measure the correlation between ground validation samples and these ground types. Unlike supervised techniques where the ground truth data cannot be used as an external accuracy measure, unsupervised classification must use the strength of the correlation as a justification for the habitat map. As with the previous techniques, accuracy measures are part of the modelling process.



The method of unsupervised accuracy measurement

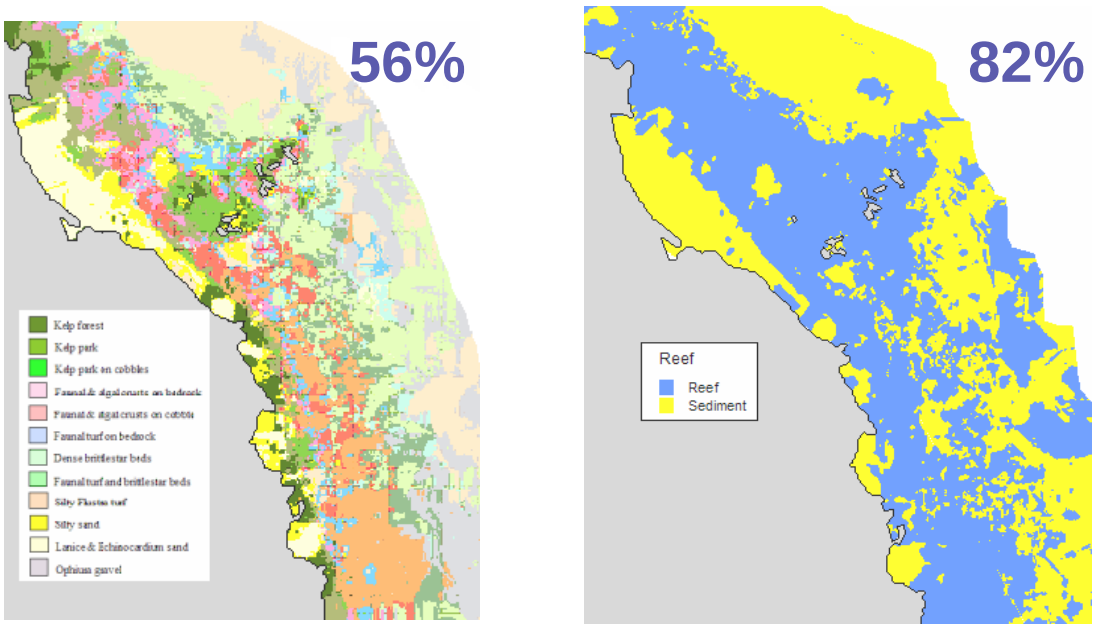


The method of supervised accuracy measurement

Although this seems like a distinct advantage over supervised techniques, the assumption that the clusters defined by the automatic process have any clear relation to the biota is questionable.

How do I interpret my accuracy assessment?

Clearly, given the complex issues surrounding accuracy, it is not a simple matter of equating accuracy with usefulness. The example below shows two versions of the same map: the data have been interpreted to 13 life form habitat classes. A number of these can be considered to be reef (with outcrops of bedrock and boulders) with varied life forms, and the remainder are sediment habitats. If the classes are amalgamated into these two groups, accuracy increases considerably. This might seem obvious. But the reason for the lower accuracy of the life form map is because there is greater confusion between, for example, kelp forest and kelp park than between the rocky habitats and the sediment habitats. It might well be the case that managers simply wish to know where reefs occur. However, the map to the left shows far more information about the distribution pattern of the component reef habitats and a degree of confusion might be acceptable especially if this is between similar habitats.



An example of how a map with fewer classes has higher accuracy, but potentially less useful to the end user

The map on the right is more accurate, but is it more useful? Accuracy is usually traded off against information content in the interpretation and analysis of a dataset. At one extreme, the interpretation may attempt to show subtle variation in habitat content that are simply not supported by the data. At the other extreme, the habitats are so general that the information is not useful for most purposes. An alternative to a measuring the accuracy of a map is to determine the confidence someone using the map for a particular purpose should have in the map. How much confidence should I put in a map? discusses how confidence can be assessed and is followed by The MESH approach to confidence assessment which describes the MESH approach to confidence assessment.

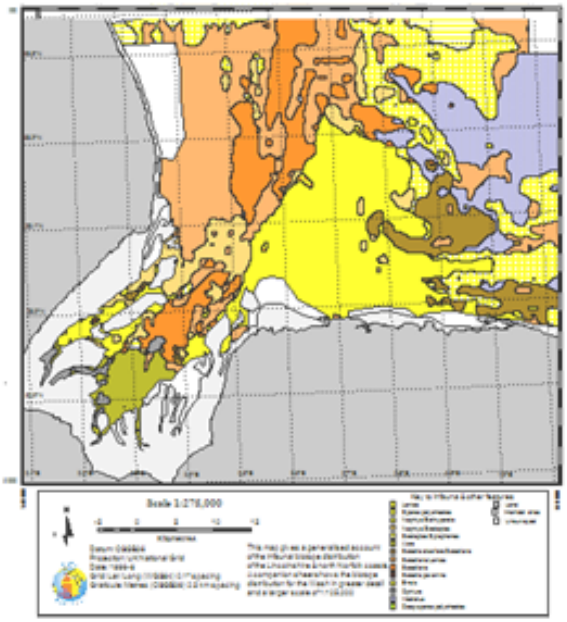
Authors: Bob Foster-Smith, Natalie Coltman, Fiona Fitzpatrick

Close

How much confidence should I put in a map?

Accuracy measures, when applied to maps from a single survey, are a valuable way for surveyors to indicate how well their map performs as a predictive tool. However it should be clear from the discussion on accuracy that there are difficulties in interpreting accuracy measures. These difficulties are compounded when habitat maps have been derived from data from many different sources. It might be possible to assess the accuracy of the contributory maps, but accuracy measurement rarely accompanies published maps and may not even have been undertaken as part of the mapping process. It may be possible to test the accuracy of final broad scale map by testing its predictive power against a test data set. However, the results may not particularly meaningful or easy to interpret.

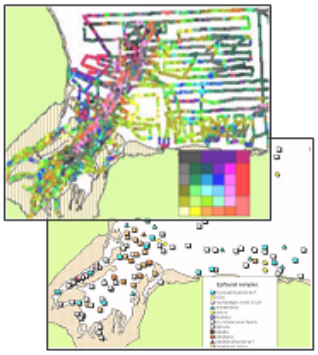
Confidence assessment might provide an alternative way of judging the usefulness of a map. Confidence is a more subjective form of assessment and is derived from a number of different criteria. This can be done by a map-user simply by checking the habitat map and any accompanying report and supporting maps for criteria that indicate the standard of mapping. This might be in the form of a check-list of questions. Does the published map show basic information about the origins of the map and its datum? Furthermore, if there is a report does it show clearly how the map was derived?



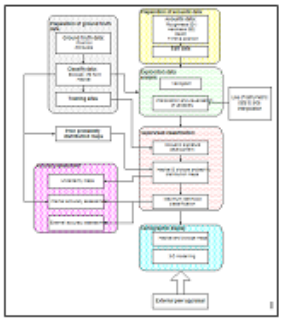
Does the map have basic information, such as:-

- Scale
- Coordinates
- Grid
- Date compiled
- Authorship
- Datum
- Legend

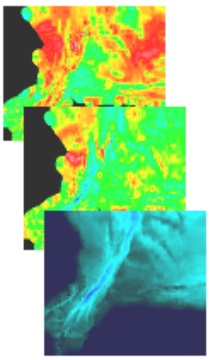
Are the source data shown?



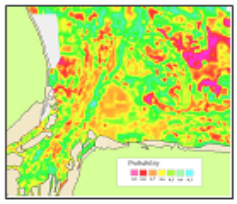
Is the process of analysis and interpretation clear?



Are critical stages of the interpretation shown?



Is there any attempt at accuracy or uncertainty measurement?



If there is a report associated with the map, does it clearly show how the map was made? For example, does the report contain some measurement of accuracy (far right image)?

Interpreted maps often show some of the supporting data: For example, the British Geological Survey's seabed sediment maps have inset maps displaying the location of grabs, cores and survey lines. The density and distribution of ground truth data gives a very good indication of the likely uncertainty associated with the interpretation. The full coverage maps of the remote sensing data should also be displayed. Often the preparation of these maps has involved some form of data manipulation. For example, AGDS point data are usually interpolated to produce a pseudo-full coverage. It is good practice to supply a map of the original track data showing the values of the data points. Interpolation can introduce spurious artefacts into the data and these are quite often obvious when comparing the point data to the interpolated data.

It is also valuable to see the process of making the habitat map as a flow chart showing the crucial stages in data manipulation and modelling. This alerts viewers to look carefully at these stages to see where errors might have been introduced into

the process. Often these supporting maps and charts are provided in the original survey reports and these may not be available to the user. If the map is being used to make decisions that are critically dependent upon the detail of a habitat map, users should try to view the full report, not just the final interpreted map.

Is there a structured way to assess confidence?

Using a simple check-list together with an inspection of any supporting maps can be further developed into a structured multi-criteria approach. Synthesising these various criteria into an overall assessment requires judgement as to the way in which scores are combined, including how different criteria should be weighted. The weightings can be used as a way of altering a confidence assessment when the purpose against which the map is being assessed changes. For example, for conservation management, information about the distribution of biological communities on the seabed is important, whereas for safe navigation this information is not important.

A systematic process has advantages over an informed but unstructured assessment in that it is transparent to others how the assessment has been made; assessments of more than one map can be compared so that, if a choice exists, the better quality map can take precedence; and, the criteria can be published so that people preparing a habitat map can ensure that the relevant data are included in anticipation of the confidence assessment.

A multi-criteria approach has been used within the MESH project and is described in detail in the section The MESH approach to confidence assessment. The approach was developed to facilitate the determination of confidence in habitat maps displayed on the MESH webGIS (http://www.searchmesh.net/webGIS). The selection of maps available includes historical maps as well as recent maps. The partnership examined and assembled the various factors that affect confidence in a map and constructed a confidence assessment methodology. The evaluation process addresses three main questions:

1. How good is the remote sensing?
2. How good is the ground truthing?
3. How good is the interpretation?

These questions were selected because MESH promotes the creation of habitat maps through the interpretation of remote sensing data and ground truthing data. The factors used to answer these questions are presented below.

Links to websites:

http://www.searchmesh.net/webGIS

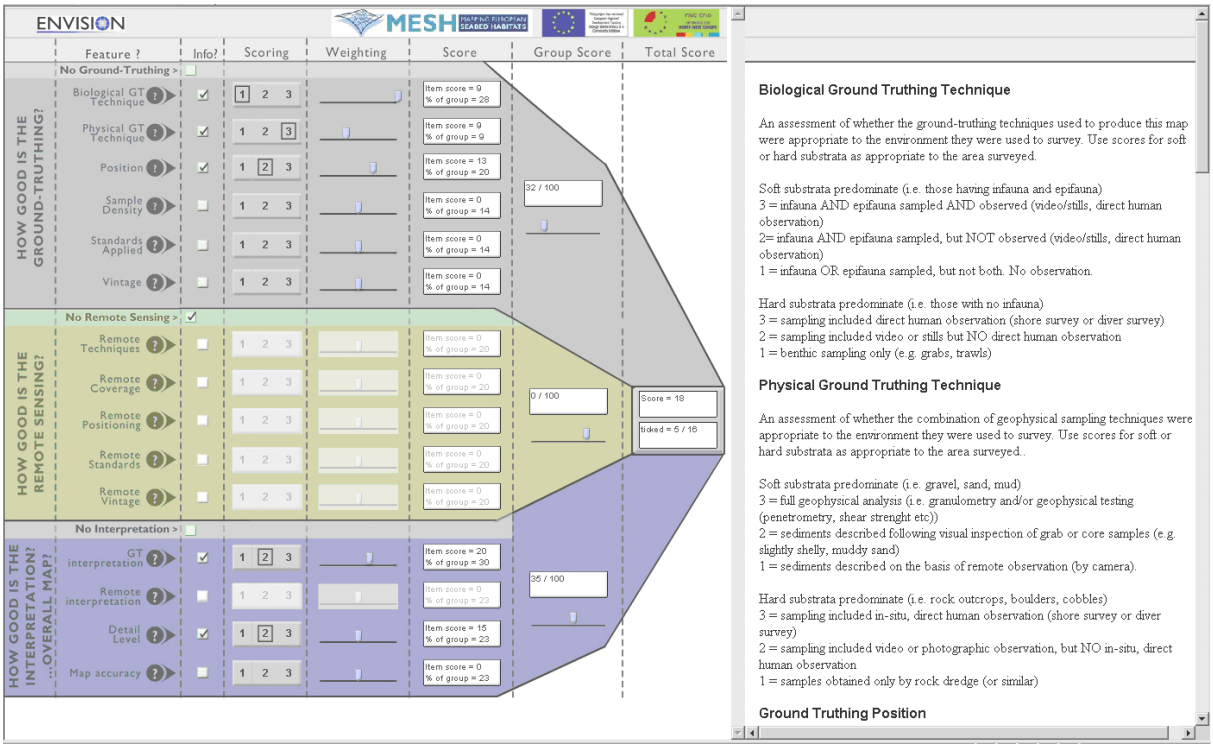
The MESH approach to confidence assessment

This section describes the systematic approach to the assessment of confidence in maps that has been developed by the MESH project. MESH has collated many maps produced for different purposes, ranging over many years, employing many different techniques and from a variety of sources. In many areas the maps overlap and users need to know the relative confidence they should have in these maps. This guidance will also be useful to future map makers to determine which factors will increase or decrease the confidence a user has in a map. The problem of confidence is multifaceted and any assessment runs the risk of being very subjective and dependant on the person undertaking the exercise. Clearly, if some comparison is to be made between maps, then the assessment should be as objective as possible.

The MESH partners decided that a confidence assessment system should be devised and the confidence factors stored as new metadata elements so that they are accessible together with discovery metadata describing the map. The metadata already compiled by the project lacked sufficient detail to make objective decisions about the confidence of various factors. The purpose of the MESH Confidence Assessment is to visualise a calculated overall confidence score on the MESH webGIS using study outlines, and to link these outlines to the full set of scores so that the assessment process remains transparent. The overall scores allow rough comparisons to be made between maps whereas the full set of scores enables users to identify why one map may have scored more highly than another. A scoring system based on a multi-criteria approach also allows survey planners to anticipate the effect of changing various survey parameters on the overall performance of a survey. In other words, it may help as a planning tool.

The MESH approach is a compromise between being comprehensive and being easy to understand and usable. Many criteria have undoubtedly been left out and the exact scores suggested for each map may be challenged. The system is *not* designed to identify subtle differences between maps, but rather to give a simple and robust assessment. The exact score for any one field could be debated, but the overall score is little affected by tweaking the individual scores for the fields. Although the way a multi-criteria scoring approach is designed and operates is open to criticism, at least it is also open to inspection because the decision points are established and guidance is given to standardise scoring carried out by different individuals.

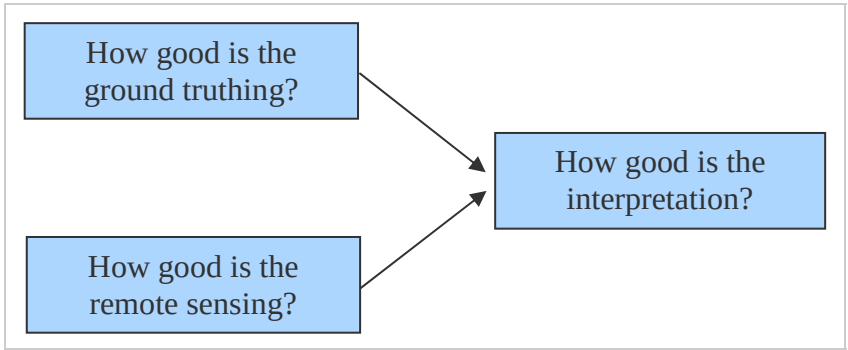
The MESH confidence assessment methodology has been built into two applications, each of which is best suited to a particular type of confidence assessment. For multiple maps, we suggest using the MS Excel Confidence Scoresheet (MESH Confidence Scoresheet.xls) for ease of data entry and comparison between maps. For a more interactive tool which is best suited to the assessment of a single map, use the MESH Confidence Tool, built as a Flash application. This tool makes it easier to see the effects of changing individual scores and weightings. For those interested in the methodology but who will not be carrying out confidence assessments, the factors and scoring system are set out in scoring guidelines MESH Confidence Assessment Guidelines. The scoring guidelines are built into each application for quick reference.



The MESH Confidence Tool, built as a Flash application. The scores and weightings of the factors are selected in the left-hand pane and the scoring guidelines are in the right-hand pane.

Design of the MESH Confidence Assessment

The design follows the general scheme for habitat mapping promoted throughout this Guide, namely that habitat mapping is essentially in three parts: ground truthing, remote sensing and combined interpretation. Thus, the criteria have been arranged in three groups:



Two of the groups refer to data collection and the third refers to data interpretation. Each group contains a number of criteria (or factors) which are scored separately and are then combined to form an overall score for each group. These three group scores are then combined into a final overall score.

- Scoring individual factors: Each is scored between 0 and 3, with the scores as follows: 0 = particular task not carried out; 1 = carried out to a low standard or carried out but to an unknown standard (and therefore assumed to be lowest

standard by default); 2 = carried out to a moderate standard; 3 = carried out to a high standard.

- Group scores: Each group score is a simple addition of all factors expressed as a percentage of the maximum score possible.
- Weightings: This means that each factor within a group contributes equally to the group score. However, this may not be considered appropriate in all cases in that some criteria may be more important than others. There is the possibility of weighting factors differently. This has been used sparingly and it is not envisaged that the weightings would normally be altered. The weightings range between 1 and 6, with the majority of weightings being 3. However, there is one area where weightings have been used to adjust relative importance of criteria: it was considered that biological ground truthing was more important than physical habitat ground truthing since habitat mapping within its use in MESH emphasises biological habitats above purely physical. For this reason the weightings for the biological ground truth criteria were given a weighting of 6 whilst the physical ground truthing criteria were given a weighting of 2. Other users may wish to adjust weightings for their particular purposes. This can be done **but any re-adjustment must be justified**.
- Overall scores: The outputs of the scoring system are scores for each group and an overall score that combines the three group scores using a simple average.

Whilst this scoring system could be made more comprehensive (by including more criteria) and more sophisticated (e.g., using a continuous scale instead of a 0-3 score), the system as it stands is simple and transparent. Experience has shown that deliberations resulting small adjustments to any particular criterion probably make little difference to the overall scores.

The scoring system was developed after the *MESH metadata catalogue* (http://www.searchMESH.net/metadata) for published maps was designed and the demands of the scoring system necessitated changes in the structure of the database. When it came to trying to assess the confidence in maps using the existing metadata fields, it was found over and over again that information was not available to assign a score to particular factors. Any information that was included in the metadata was given in free text format which would have made an assessment very subjective. To overcome this obstacle it was decided that a transparent numeric scoring system was the key to confidence assessment. Thus a new 'set' of metadata fields for assessing the confidence in maps was developed. These have been added to the MESH metadata standard (as another tab in the spreadsheet, imported into the database for the MESH metadata catalogue as a linked table). A confidence assessment is being carried out for each entry relating to a map of the seabed,

In MESH, the collation of seabed mapping data means there are overlapping seabed maps in some areas: for the purpose of creating a single layer of translated maps it is necessary to decide which of these maps to use in areas where there are overlaps. One of the first applications of the MESH confidence assessment was to maps which overlapped and to order these in terms of their confidence scores.

Selection and scoring of confidence factors

A brief description of each factor is given below in the three groups of remote sensing, ground truthing and data interpretation. Each group is followed by a table presenting scoring guidelines for the factors.

Data collection: Remote sensing factors

The following factors were selected in order to answer the question *How good is the remote sensing data collection?*

Remote technique

It is not practical to develop a '1 to 3' scoring system to cover all possible acoustic techniques and combinations of techniques. A more pragmatic approach is to use your own judgement of whether the technique(s) used in the remote sensing survey were appropriate to distinguish between the expected ground types in the area.

Remote coverage

This score has two aspects: confidence in remote sensing data will be higher if the coverage is better (ideally overlapping data to provide 'replicates'), and; confidence in remote sensing data from a homogenous area will be higher than if the area surveyed was heterogeneous. Therefore the MESH system takes account of both these aspects so that, for example, wider track spacing of AGDS is more acceptable for homogenous areas than for heterogeneous areas, which will require narrower track spacing to obtain the same confidence score.

Remote positioning

The positioning system is used here as a proxy for the precision of the positioning when collecting the remotely sensed data, because different systems will have different ranges of precision. The remote sensing data may have been collected using a different positioning system than was used to collect the ground truth data, so there is a score for Remote positioning and Ground truth positioning.

Remote standards applied

Following accepted standards during data collection gives an indication of the quality of the data. The standards used can be externally accepted (highest score), or internal to the organisation collecting the data (lower score). Data collected to internal standards score more highly than those which lack clear standards of data collection.

Remote vintage

The age (vintage) of the remote sensing data indicates the likelihood of change occurring on the seabed between the time the data were collected and the present day. It was not practical in this system to include an assessment of the environmental variability, including human impact, but it should be remembered that some habitats are temporally variable whereas others are static on a decadal time-scale (compare sand wave fields and bedrock outcrops). This issue is further complicated because a map may include both variable and static habitats.

How good is the remote sensing			Score	Guidance
Remote technique	Were the techniques used appropriate for the ground type?	An assessment of whether the remote technique(s) used to produce this map were appropriate to the environment they were used to survey.*	3	technique(s) highly appropriate
			2	technique(s) moderately appropriate
			1	technique(s) not appropriate
*N.B. If necessary, adjust your assessment to account for technique(s) which, although appropriate, were used in deep water and consequently have a significantly reduced resolution (i.e size of footprint):				
Remote coverage	Was the ground covered appropriately?	An assessment of the coverage of the remote sensing data including consideration of heterogeneity of the seabed: (See Coverage x Heterogeneity matrix below)	Coverage scores – use these to determine coverage then	
			3	good coverage; 100% (or greater) coverage or AGDS track spacing <50m
			2	moderate coverage; swath approx 50% coverage or AGDS track spacing <100m
			1	poor coverage; large gaps between swaths or AGDS track spacing >100m
			Final scores	
			3	good coverage OR moderate coverage + low heterogeneity
			2	moderate coverage + moderate heterogeneity OR poor coverage + low heterogeneity
			1	moderate coverage + high heterogeneity OR poor coverage + moderate or high heterogeneity
Remote positioning	How were the positions determined for the remote data?	An indication of the positioning method used for the remote data:	3	differential GPS
			2	GPS (not differential) or other non-satellite 'electronic' navigation system
			1	chart based navigation, or dead-reckoning
Remote standards applied	Were standards applied to the collection of the remote data?		3	remote data collected to approved standards
			2	remote data collected to 'internal' standards
			1	no standards applied to the collection of the remote data
Remote vintage	How recent are the remote data?	An indication of the age of the remote data:	3	< 5yrs old
			2	5 to 10 yrs old
			1	> 10 yrs old

Heterogeneity scores	
3	low; habitats form homogeneous patches > 100x100m
2	moderate; habitats patches between 50 x 50m and 100 x 100m
1	high; habitat patches regularly < 50 x 50m
Coverage scores	
3	good coverage; 100% (or greater) coverage or AGDS track spacing <50m
2	moderate coverage; swath approx 50% coverage or AGDS track spacing <100m
1	poor coverage; large gaps between swaths or AGDS track spacing >100m

		Heterogeneity		
		Low	Moderate	High
Covergare	Poor	2	1	1
	Moderate	3	2	1
	Good	3	3	2

The heterogeneity/coverage matrix above should be used to derive scores for Remote coverage. First assess scores for each of the two components using the table on the left, and then find the appropriate overall Remote coverage score from the matrix on the right. Note that this separate heterogeneity/coverage process is a rough guide; the score is only one of many that will contribute to the overall score.

Data collection: Ground truthing factors

The following factors were selected in order to answer the question *How good is the ground truthing data collection*?

Close

Biological ground truth technique

It is not practical to develop a '1 to 3' scoring system to cover all possible biological ground truthing techniques and combinations of techniques. However, scoring guidelines are provided for combinations of types of biological ground truth data (for example, human observation, video, or benthic sampling). The most appropriate combination will differ depending on the substrate; scores in the table below show the combinations for ground truth data collected where hard substrata dominate and where soft substrata dominate. This factor was selected to highlight that biological ground truth data is very important in habitat mapping for conservation management, and consequently this factor is weighted more heavily than physical ground truth technique.

Physical ground truth technique

As for biological ground truth technique, it is not practical to develop a '1 to 3' scoring system to cover all possible physical ground truthing techniques and combinations of techniques. However, scoring guidelines are provided for combinations of types of physical ground truth data (for example, human observation, video, or benthic sampling). The most appropriate combination will differ depending on the substrate; scores in the table below show the combinations for ground truth data collected where hard substrata dominate and where soft substrata dominate. Physical ground truthing can still help to create a seabed map in cases it was not appropriate to carry out biological sampling; this factor has a lower weighting than Biological ground truth technique

Ground truth positioning

The positioning system is used here as a proxy for the precision of the positioning, because different systems will have different ranges of precision. The ground truth data may have been collected using a different positioning system than was used to collect the remote sensing data, so there is a score for Ground truth positioning and Remote positioning.

Ground truth density

The number of times each class in the map was sampled will affect the confidence in the map; more ground truth samples in a class means that (where they agree with each other and the class!) those samples are more likely to be a good representation of that class.

Ground truth standards applied

Following accepted standards during data collection gives an indication of the quality of the data. The standards used can be externally accepted (highest score), or internal to the organisation collecting the data (lower score). Data collected to internal standards score more highly than those which lack clear standards of data collection.

Ground truth vintage

The age (vintage) of the ground truth data indicates the likelihood of change occurring on the seabed between the time the data were collected and the present day. It was not practical in this system to include an assessment of the environmental variability, including human impact, but it should be remembered that some habitats are temporally variable whereas others are static on a decadal time-scale (compare

sand wave fields and bedrock outcrops). This issue is further complicated because a map may include both variable and static habitats.

How good is the ground truthing?			Score	Guidance
Biological ground truthing	Were the techniques used appropriate for the habitats encountered?	An assessment of whether the ground-truthing techniques used to produce this map were appropriate to the environment they were used to survey. Use scores for soft or hard substrata as appropriate to the area surveyed.	Soft substrata (infauna and possibly epifauna)	
			3	Infauna AND epifauna sampled AND observed (video/stills, direct human observation)
			2	infauna AND epifauna sampled, but NOT observed (video/stills, direct human observation)
			1	infauna OR epifauna sampled, but not both. No observation.
			Hard substrata (infauna not significant)	
			3	sampling included direct human observation (shore survey or diver survey)
			2	sampling included video or stills but NO direct human observation
			1	benthic sampling only (e.g. grabs, trawls)
Physical ground truthing	How appropriate were the sampling techniques to determining the geophysical nature of the seabed?	An assessment of whether the combination of geophysical sampling techniques was appropriate to the environment they were used to survey. Use scores for soft or hard substrata as appropriate to the area surveyed.	Soft substrata predominate (gravel, sand, mud)	
			3	full geophysical analysis: granulometry and/or geophysical testing (e.g. penetrometry, shear strength)
			2	sediments described following visual inspection of grab or core samples (e.g. slightly shelly, muddy sand)
			1	sediments described on the basis of remote observation (by camera).
			Hard substrata predominate (rock outcrops, boulders,	
			3	sampling included in-situ, direct human observation (shore survey or diver survey)
			2	sampling included video or photographic observation, but NO in-situ, direct human observation
			1	samples obtained only by rock dredge (or similar)
Ground truth positioning	How were the positions determined for the ground-	An indication of the positioning method used for the ground-truth data:	3	differential GPS
			2	GPS (not differential) or other non-satellite 'electronic' navigation system
			1	chart based navigation, or dead-reckoning
Ground truth density	Was the density of sampling adequate?	An assessment of what proportion of the polygons or classes (groups of polygons with the same 'habitat' attribute) actually contain ground-truth data:	3	Every class in the map classification was sampled at least 3 times
			2	Every class in the map classification was sampled
			1	Not all classes in the map classification were sampled (some classes have no ground-truth data)
Ground truth standards applied	Were standards applied to the collection of the ground-truth data?	An assessment of whether standards have been applied to the collection of the ground-truth data. This field gives an indication of whether some data quality	3	ground-truth samples collected to approved standards
			2	ground-truth samples collected to 'internal' standards
			1	no standards applied to the collection of ground-truth samples
Ground truth vintage	How recent are the ground-truth data?	An indication of the age of the ground-truth data:	3	< 5yrs old
			2	5 to 10 yrs old
			1	> 10 yrs old

Interpretation factors

The following factors were selected in order to answer the question *How good is the data interpretation*?

Ground truth interpretation

Expert taxonomy is important in interpreting ground truth samples. Note that maps made from only physical (rather than biological) ground truth data will score a maximum of 2 because the report does not include taxon lists.

Remote interpretation

There are a wide range of interpretation techniques and combinations of techniques which could be used to draw the polygon outlines in the habitat map. These techniques will vary depending on the type of data collected. Use your judgement to determine whether the interpretation technique is appropriate for the particular type of remote sensing data collected. Documenting the methods of interpretation used is vital: maps will get a higher score if the report provides documentation of the interpretation of the remote sensing data.

Level of detail

This factor provides an indication of the level of information included in the interpreted map. Although there may be more certainty that a class with a low level of detail has been correctly assigned, higher levels of detail are necessary for conservation management. Maps with low levels of detail be less useful for this purpose and will therefore get a lower score in this factor (note that MESH assesses the confidence of maps for their use in conservation management).

Map accuracy

A formal test of the accuracy of a map is an important component of assessing the confidence in a map. Remember that accuracy as applied to habitat mapping is a measure of the predictive power of a map to represent the world as measured against reality and error is a measure of the departure of a map from reality. It is a mathematical measure based on 'hits and misses' (successful predictions and erroneous predictions). External accuracy assessments are rare in marine habitat mapping but provide an extremely valuable test of the accuracy of a map.

How good is the data interpretation?			Score	Guidance
Ground truth interpretation	How were the ground-truthing data interpreted?	An indication of the confidence in the interpretation of the ground-truthing data. Score a maximum of 1 if physical ground-truth data but no biological ground-truth data were collected:	3	Evidence of expert interpretation; full descriptions and taxon list provided for each habitat class
			2	Evidence of expert interpretation, but no detailed description or taxon list supplied for each habitat class
			1	No evidence of expert interpretation; limited descriptions available
Remote interpretation	Were the remote data appropriately interpreted?	An indication of the confidence in the interpretation of the remotely sensed data. Note that interpretation techniques can range from 'by eye' digitising of side scan by experts to statistical classification techniques.	3	Appropriate technique used and documentation provided
			2	Appropriate technique used but no documentation provided
			1	Inappropriate technique used
Level of detail	What level of information is contained?	The level of detail to which the 'habitat' classes in the map have been classified:	3	Classes defined on the basis of detailed biological analysis
			2	Classes defined on the basis of major characterising species or lifeforms
			1	Classes defined on the basis of physical information, or broad biological zones
Map accuracy	How accurate is the map at representing reality?	A test of the accuracy of the map:	3	high accuracy, proven by external accuracy assessment
			2	high accuracy, proven by internal accuracy assessment
			1	ow accuracy, proved by either external or internal assessment OR no accuracy assessment made

Authors: Bob Foster-Smith, Natalie Coltman, Fiona Fitzpatrick

Links to resources:

MESH Confidence Scoresheet.xls

MESH Confidence Tool

MESH Confidence Assessment Guidelines

MESH Confidence Scoresheet and Tool

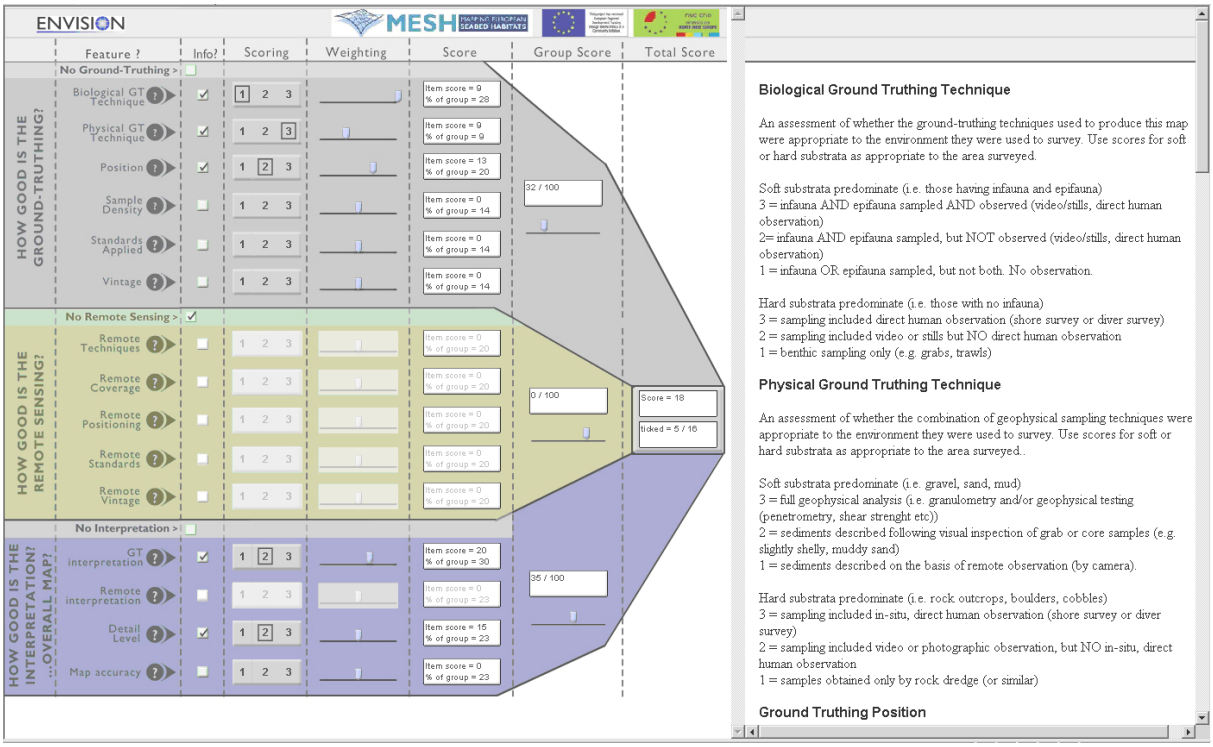
The MESH confidence assessment methodology has been built into two applications, each of which is best suited to a particular type of confidence assessment. For ease of data entry and comparison between multiple maps, we suggest using the MESH Confidence Scoresheet, an MS Excel workbook MESH Confidence Scoresheet.xls. For a more interactive tool which is best suited to the assessment of a single map, use the MESH Confidence Tool (http://www.searchMESH.net/confidence), built as a Flash application. This tool makes it easier to see the effects of changing individual scores and weightings. For those interested in the methodology but who will not be carrying out confidence assessments, the factors and scoring system are set out in the scoring guidelines (MESH Confidence Assessment Guidelines.doc).The scoring guidelines are built into each application for quick reference. The MS Excel workbook and Flash application perform the scoring task in the same way.

The workbook consists of the main sheet for entering factor scores, a weightings sheet (which performs some calculations whose results are returned to the scoresheet) and the scoring guidelines. Although the weightings sheet is available to view, it is *strongly recommended* that the weightings are not edited. For maps assessed by the MESH Project, a standard set of weightings is used for all maps. The scoring system is illustrated here using two examples. One (A) is from Sussex and this example is also referred to in How do I interpret my confidence assessment?. This is an 'old' survey using only AGDS and drop down video. The tracking was done with irregularly but generally widely spaced tracks. The surveys were undertaken over three years with very limited time and resources available. Thus, the survey should not expect to rate very highly in terms of confidence. The second example (B) is a recent survey of the Moray Firth using an interferometric swath system (bathymetry and side scan backscatter) and AGDS, combined together with towed video and grab sampling as ground truthing techniques. The coverage was 100% (swath) with a track spacing of 75-150m. Thus, it might be expected that this survey would have a higher confidence score.

	How good is the remote sensing?					**How good is the ground truthing?**						**How good is the interpretation?**				**Summary**			
Survey	RemoteTechnique	RemoteCoverage	RemotePositioning	RemoteStdsApplied	RemoteVintage	BGTTechnique	PGTTechnique	GTPositioning	GTDensity	GTStdsApplied	GTVintage	GTInterpretation	RemoteInterpretation	DetailLevel	MapAccuracy	**Overall % score**	GT % score	Remote % score	Interpretation % score
A - Sussex	1	1	2	1	1	2	1	2	2	2	1	2	2	2	2	**55**	58	40	67
B - Moray Firth	3	3	3	3	3	3	2	3	2	3	3	3	3	2	2	**92**	92	100	83

Close

The Flash application is more interactive and has the scoring guidelines available for assistance on the right-hand pane in the window. In the version of the Flash application provided on the MESH website, the weightings column is not editable.



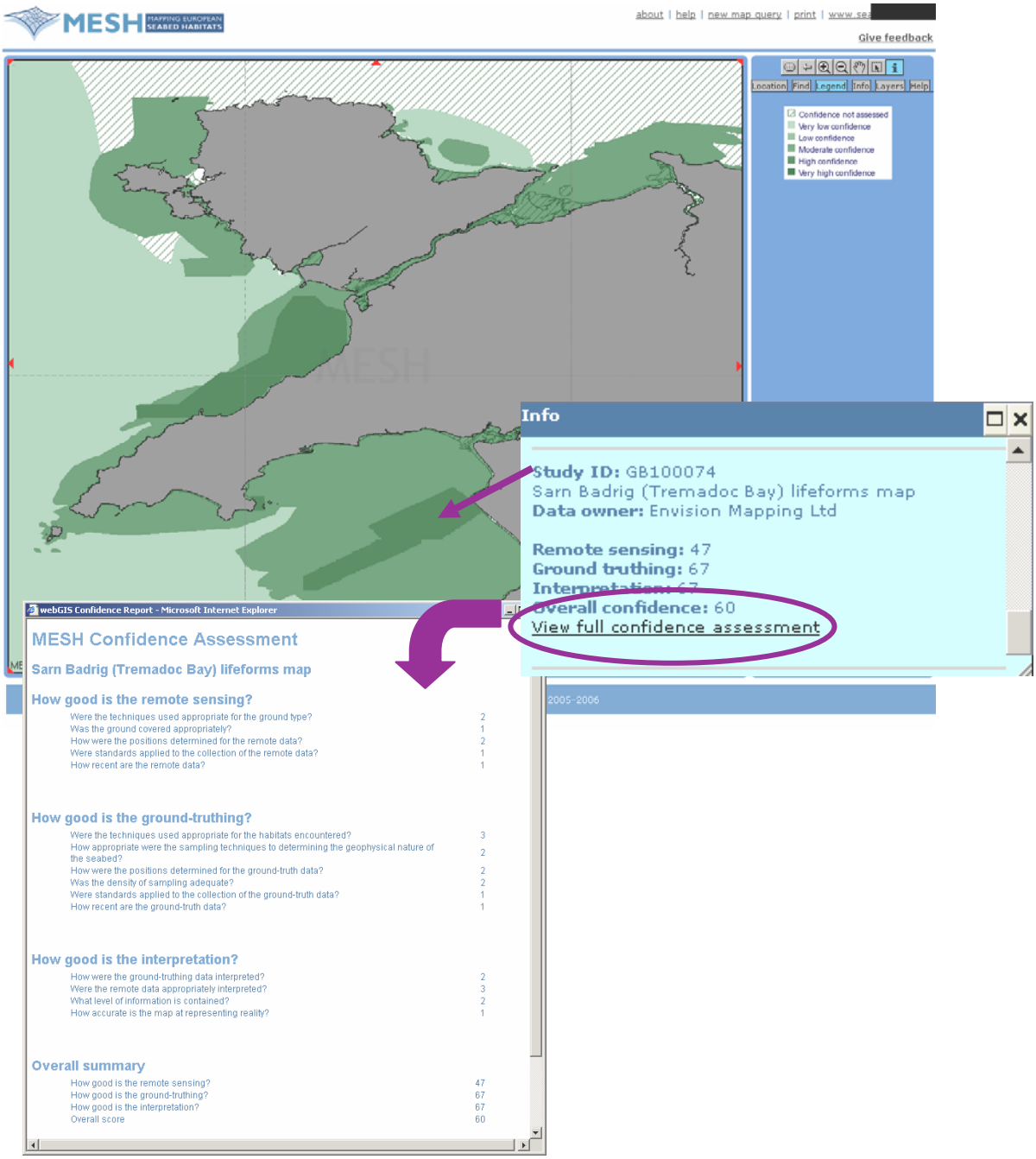
The MESH Confidence Tool was built as a Flash application. The scores and weightings of the factors are selected in the left-hand pane and the scoring guidelines are in the right-hand pane.

Displaying confidence on the MESH webGIS

These confidence scores are displayed on the *MESH webGIS* linked to outlines of the study areas. MESH uses a colour ramp with five grades of increasing intensity. The cut-offs were chosen so that maps must meet certain minimum requirements in order to get into each group. The example below is from the surveys off the coast of north Wales. One drawback of the scheme at present is that the darker colours (surveys with a high confidence) obscure the lighter colours (low confidence surveys). However, the implication of the scheme is that viewers are given an indication of which of the overlapping surveys should be accorded with a higher confidence.

WARNING! The confidence scores must be used with caution. Their purpose is to give guidance to viewers and help them come to some conclusion about the usefulness of the available maps. Ultimately the use of a map can only be assessed given a proper understanding of the particular purpose the viewer has in mind. It is up to map users to make their own assessment. However, prioritising available maps may be done through the scoring system together with the metadata. It is then up to the user to obtain the map and (if possible) the associated reports and make the final decision as to whether they rely on the information in the map.

Close



An example of how the confidence scores are displayed on the MESH webGIS; the summary scores are presented in an 'Info window' with the detailed scores available through a linked web page.

Links to resources:

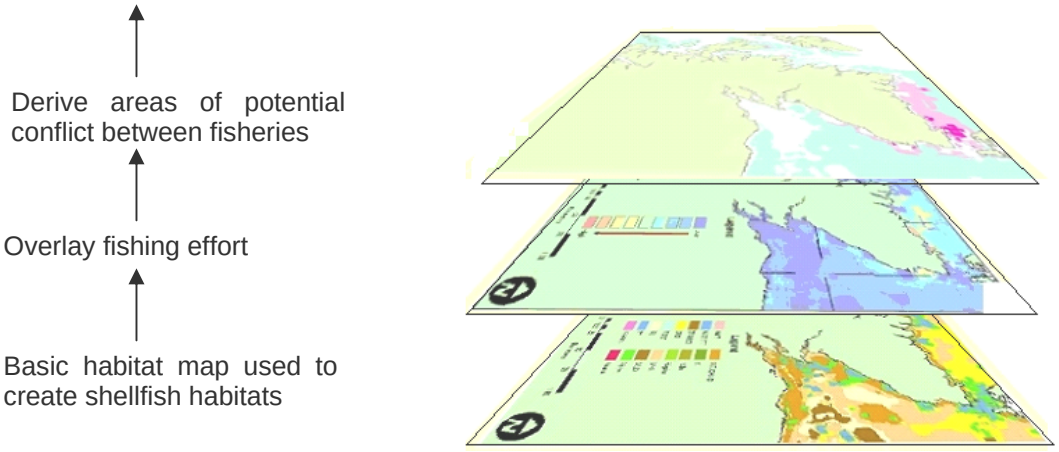
MESH Confidence Scoresheet.xls

Confidence_Tool\confidenceAssessment.html

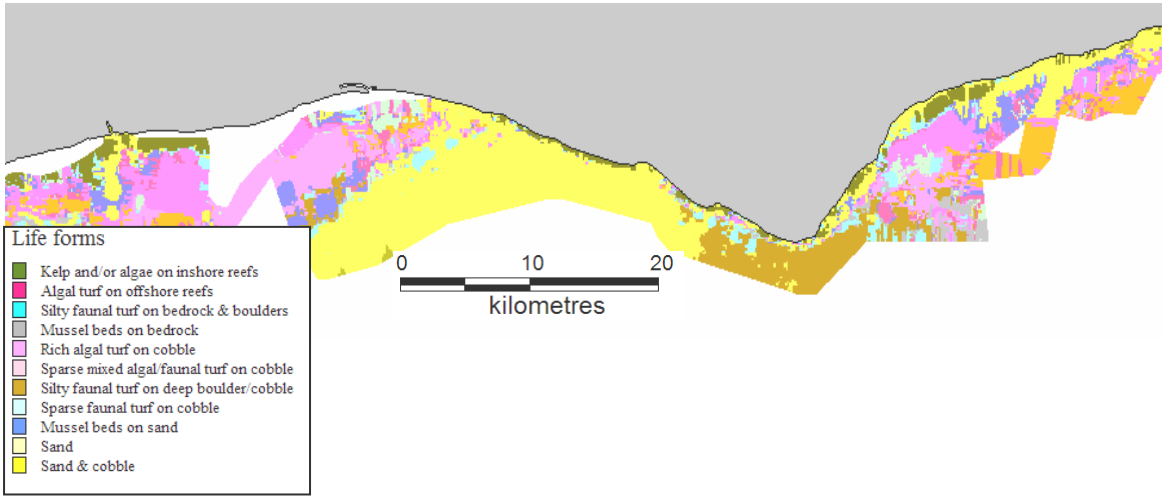
MESH Confidence Assessment Guidelines.doc

How do I interpret my confidence assessment?

All habitat maps are predictive and can only really be tested through usage. Maps attract more confidence if they have been inspected and approved by external experts and those with local knowledge. Maps may have been used and found to predict well using independent validation. They might, more generally, have been used by stakeholders and found to be acceptable and stood the test of time.



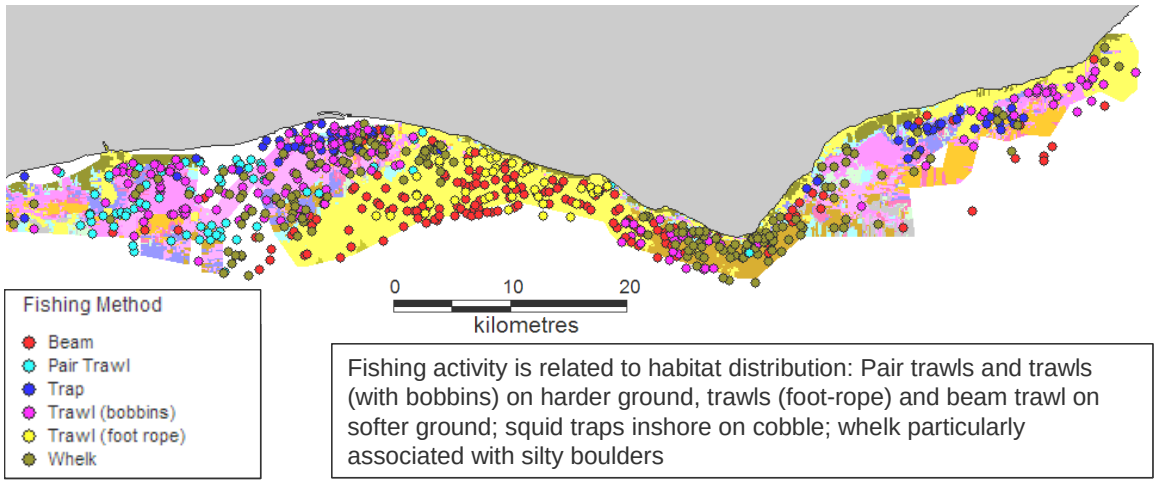
Maps may not score highly on any measure of confidence. But this does not necessarily mean that the have little use in certain applications. Indeed, such a map may turn out to match well to a completely independent data source and this can lead us to revise our opinion of a map. For example, the habitat map below was derived from data that were collected between 1994 and 1996, on different vessels, using only AGDS, with varied track spacing (often more than 500m apart), with an uneven spread of ground truth data collected using only towed video. Using the scoring system, the map attains an overall score of 51%, which is quite a low confidence score.



The second map has overlain fisheries sightings data (coded according to fishing method). These data, collected between 2004 and 2006 also have their limitations as

Authors: Bob Foster-Smith, Natalie Coltman, Fiona Fitzpatrick

to accuracy. However, the correspondence between fishing activity and habitat distribution is readily apparent.



Fishing activity is related to habitat distribution: Pair trawls and trawls (with bobbins) on harder ground, trawls (foot-rope) and beam trawl on softer ground; squid traps inshore on cobble; whelk particularly associated with silty boulders

Thus, maps may prove to be useful even if they do not attract high confidence initially. It also follows that the survey strategy might also not be ideal, but all that is possible given the constraints, but still produce maps of some value.

Authors: Bob Foster-Smith, Natalie Coltman, Fiona Fitzpatrick

Acknowledgements

The MESH Partnership wishes to acknowledge the efforts of the Accuracy and Confidence Working Group in the compilation of the tables and the considerable work carried out by Dan Foster-Smith in the creation of the Confidence Tool application.